



Cost-Sensitive Uncertainty Hypergraph Learning for Identification of Lymph Node Involvement With CT Imaging

Qianli Ma^{1*}, Jielong Yan², Jun Zhang³, Qiduo Yu¹, Yue Zhao¹, Chaoyang Liang¹ and Donglin Di^{2*}

¹ Department of Thoracic Surgery, China-Japan Friendship Hospital, Beijing, China, ² The School of Software, Tsinghua University, Beijing, China, ³ Tencent AI Lab, Shenzhen, China

OPEN ACCESS

Edited by:

Kuanquan Wang,
Harbin Institute of Technology, China

Reviewed by:

Yuyao Zhang,
ShanghaiTech University, China
Li Wang,
University of North Carolina at
Chapel Hill, United States

*Correspondence:

Qianli Ma
mars_qhma@163.com
Donglin Di
donglin.ddl@gmail.com

Specialty section:

This article was submitted to
Precision Medicine,
a section of the journal
Frontiers in Medicine

Received: 21 December 2021

Accepted: 17 January 2022

Published: 10 February 2022

Citation:

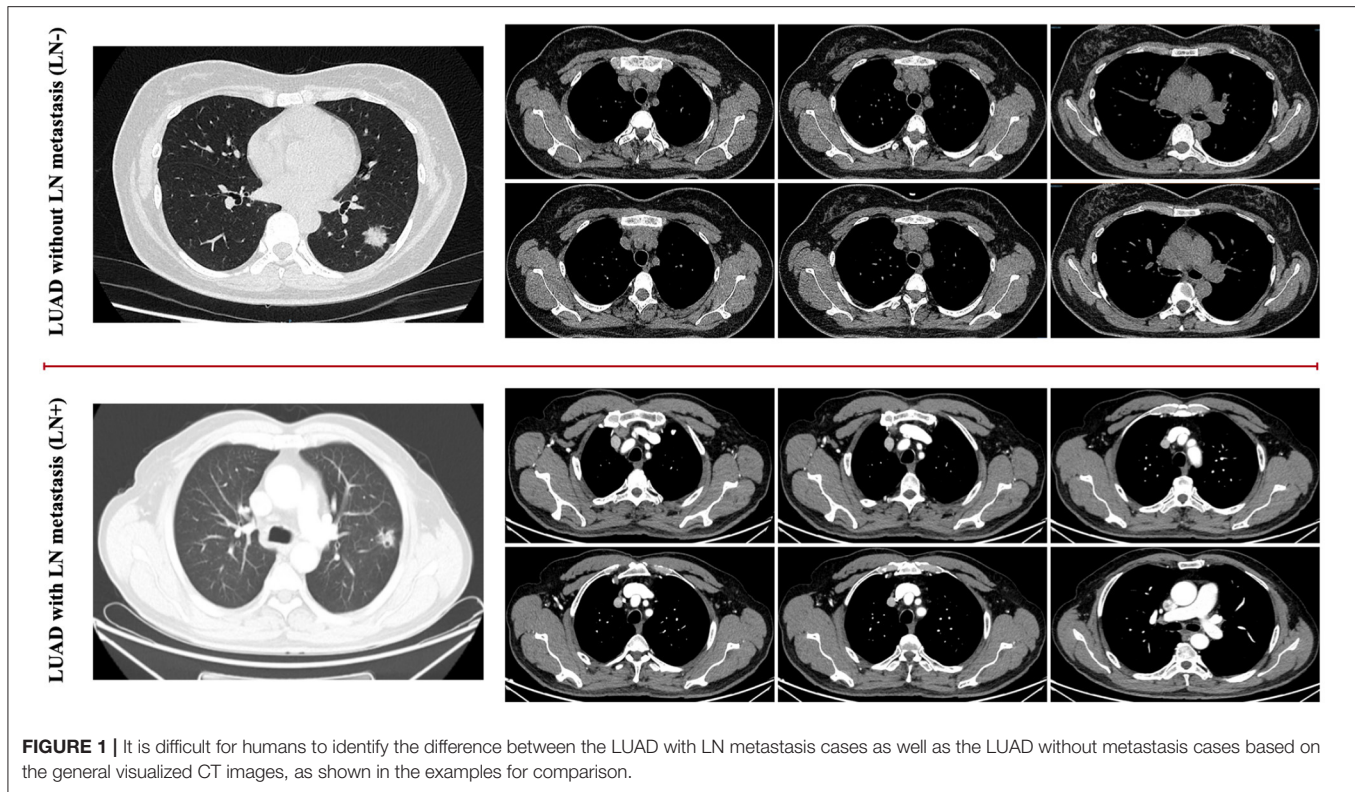
Ma Q, Yan J, Zhang J, Yu Q, Zhao Y,
Liang C and Di D (2022)
Cost-Sensitive Uncertainty
Hypergraph Learning for Identification
of Lymph Node Involvement With CT
Imaging. *Front. Med.* 9:840319.
doi: 10.3389/fmed.2022.840319

Lung adenocarcinoma (LUAD) is the most common type of lung cancer. Accurate identification of lymph node (LN) involvement in patients with LUAD is crucial for prognosis and making decisions of the treatment strategy. CT imaging has been used as a tool to identify lymph node involvement. To tackle the shortage of high-quality data and improve the sensitivity of diagnosis, we propose a Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL) model to identify the lymph node based on the CT images. We design a step named “Multi-Uncertainty Measurement” to measure the epistemic and the aleatoric uncertainty, respectively. Given the two types of attentional uncertainty weights, we further propose a cost-sensitive hypergraph learning to boost the sensitivity of diagnosing, targeting task-driven optimization of the clinical scenarios. Extensive qualitative and quantitative experiments on the real clinical dataset demonstrate our method is capable of accurately identifying the lymph node and outperforming state-of-the-art methods across the board.

Keywords: lymph node involvement, CT imaging, hypergraph learning, cost-sensitive, lung cancer

1. INTRODUCTION

Lung cancer is the most commonly diagnosed cancer and the leading cause of cancer death worldwide (1, 2). About 2.1 million new lung cancer cases and 1.8 million deaths were predicted in 2018 (3). In 2020, these two numbers rise to 2.2 million and 1.8 million, respectively, (2). Lung adenocarcinoma (LUAD) is the most common type of lung cancer (4–6). The presence of metastasis in the lymph nodes (7) is an important prognostic factor in lung cancer. Accurate identification of lymph node (LN) involvement in patients with LUAD, as shown in **Figure 1**, is crucial for prognosis and treatment strategy decisions (8, 9). Patients without metastatic lymph nodes, or with only intrapulmonary or hilar lymph nodes, are generally considered candidates for straightforward resection. Although the sub-types of LUAD are found related to the predictors of LN metastasis, they are available postoperatively (10). Information of the preoperative LN metastasis is valuable for the adequacy of surgical resection and the decision of the adjuvant therapy (11). The accurate prediction of pathologic stage for patients with lung cancer is of utmost importance. Pathologic tumor stage is considered a pivotal factor relating to survival in NSCLC, and the 5-year survival rates vary from 83% in pathological stage IA to 23% in stage IIIA tumors (12). Computed tomography (CT) is commonly used for the evaluation of pulmonary



nodules (13, 14). Many studies were designed to determine whether pulmonary nodules are benign or malignant. Zhong et al. (15) propose to use relief-based feature method and support vector machine to evaluate the impact of radiomics features in predicting the prognosis of occult mediastinal lymph node metastasis in lung adenocarcinoma. Dai et al. (16) find that lymph node micrometastases are more frequently seen in adenocarcinomas with a micropapillary component, which could give suggestive prognostic information to patients with stage I resected lung adenocarcinoma with a micropapillary component. CT has been widely used as a noninvasive diagnostic modality for diagnosis, clinical staging, survival prediction, surveillance of therapeutic response in lung cancer patients (17, 18). However, few studies have used chest CT to explore whether lymph node metastasis in LUAD (19). Therefore, in order to make a better decision on the prognosis and treatment strategies of lung cancer, as well as more fully grasp the information of lymph nodes, in this work, we utilize CT to predict whether LUAD has lymph node metastasis.

There are two main challenges of identifying the lymph node with CT imaging, listed below, that motivate our approach.

1. Noisy data, due to the collection of clinical CT images using different reconstruction kernels and CT manufacturers, along with possible patient movements;
2. The reliable sensitivity of diagnosis is relatively more important and meaningful than other criteria in the clinical scenario.

For the first challenge, a few current research works are proposed to tackle the issue of clinical data quality, mainly focusing on noise and artifact reduction, super resolution and other aspects (20). Zhang and Yu (21) propose to train their convolutional neural network using virtual metal-inserted CT images, targeting on the noise of metal artifacts. Tan et al. (22) further utilizes the SRGAN neural network to reconstruct super resolution images from the original chest CT images to improve the resolution and ultimately improve the classification results of COVID-19. Due to the scale of available data in this task being limited, we adopt the two uncertainty measurements (23) to improve the quality of pathological representations, *i.e.*, epistemic and the aleatoric uncertainty, respectively, generated by the “Classifier Measuring” and “Statistical Measuring.” In this manner, our model is capable of allocating the different attentional weights combined with the two uncertainty measurements. Due to ignoring the underlying correlation between samples, some machine learning methods such as Random forest, Boosting, or CNN are lacking in effect, but graph learning and further hypergraph learning methods can make up for this deficiency. Hypergraph Learning methods (24–26) perform well on generate the high-order representations for complex data, such as whole-slide images (WSI) (27), CT imaging (23), drug-target interactions (28), *etc.* Therefore, given the data with uncertainty weights, we further propose an uncertainty hypergraph learning to extract the high-order representations from the CT images, which augments the pathological informative features effectively.

With regards to the second challenge, several works have made efforts on lymph node involvement. Zhou et al. (29) studied one or a combination of machine learning methods in Logistic regression, Random forest, XGBoost, and GBDT to construct lymph node metastasis in patients with poorly differentiated intramucosal gastric cancer. Supervised machine learning methods including random forest classifier, artificial neural network, decision tree, gradient boosting decision tree, extreme gradient boosting, and adaptive boosting can also be used to predict central lymph node metastasis in patients with papillary thyroid cancer (30). Besides improving the accuracy of overall prediction, we further focus on boosting the performance on sensitivity by the designed “Two-Stage Cost Sensitive Hypergraph Learning.” One stage is to capture the cost sensitivity of negative cases in the latent feature spaces, which will enable the hypergraph model to allocate the lymph node involvement cases more weights. The other is called “Supervising Cost Sensitivity,” making the loss function supervise the hypergraph model with more attached importance on the patients with LUAD for individual preoperative prediction of LN metastasis. Combining the structures introduced above, the overall framework we proposed, named as “Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL)” has the ability to identify the lymph node accurately and outperform state-of-the-art methods on our collected real clinical dataset.

The main contributions of this paper are summarized as follows:

1. We propose a framework—CSUHL to tackle the task of identifying the lymph node, focusing on the uncertainty measurement of clinical CT imaging, as well as the cost sensitive hypergraph learning for identifying.

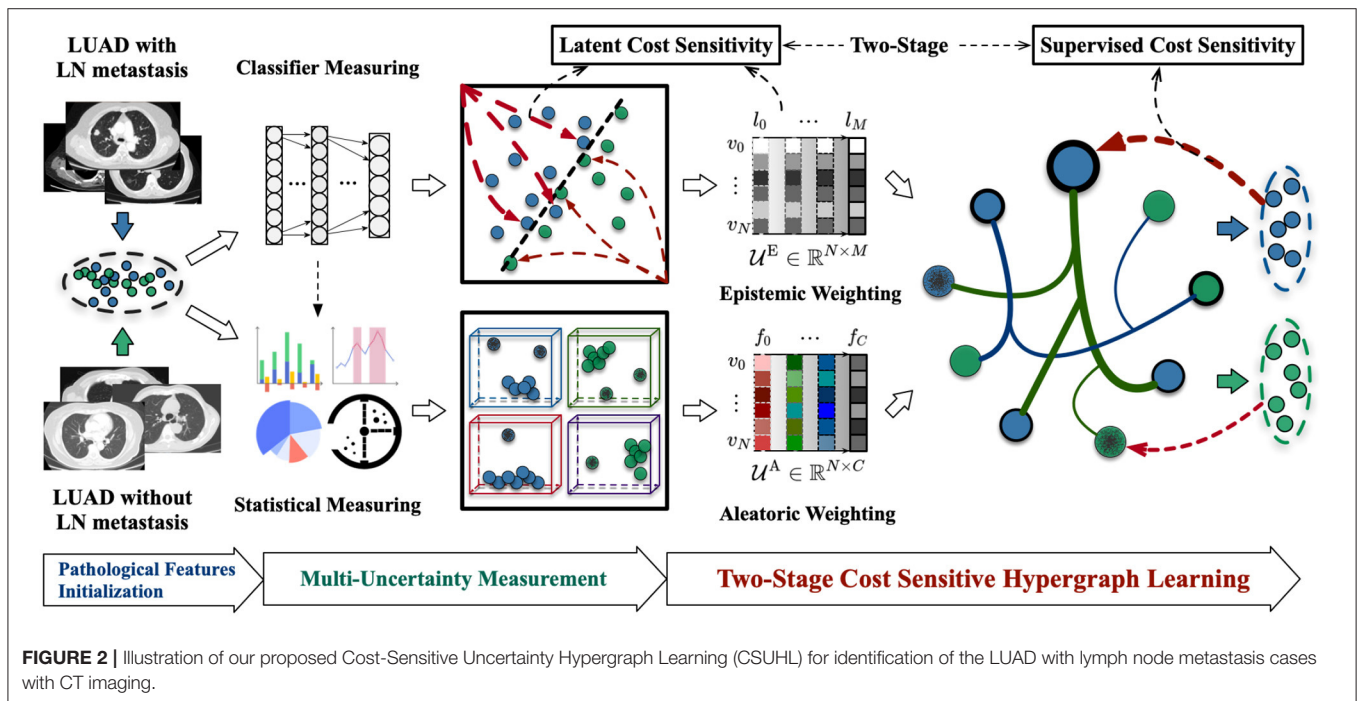
2. On our collected real clinical dataset, we conduct extensive experiments to demonstrate the proposed method consistently outperforms state-of-the-art methods across-the-board, relatively improving the performance on the accuracy (ACC), sensitivity (SEN), specificity (SPEC), Balance (BAC) by up to 3.90, 8.00, 2.02, and 4.95%, respectively, compared with the previous best method.

2. MATERIALS AND METHODS

In this section, we will first introduce the materials we collected and the processing for extracting the initial features in details. Then we will illustrate our proposed framework—“Cost-Sensitive Uncertainty Hypergraph Learning (CSUHL),” as shown in **Figure 2**, which is composed of three steps, *i.e.*, “Pathological Features Initialization,” “Multi-Uncertainty Measurement,” and “Two-Stage Cost Sensitive Hypergraph Learning,” respectively.

2.1. Materials and Preprocessing

In this study, a total of 61 CT images were collected, including 35 from lymph node negative patients and the rest 26 from lymph node positive patients. These images were provided by the China-Japan Friendship Hospital. All the cases were acquired from January 2017 to March 2019. The CT scanners used in this study include Aquilion ONE from TOSHIBA, MEDICAL System Revolution from GE, and SOMATOM Definition Flash from SEMENS. The CT protocol here includes: 120KV, reconstructed CT thickness is 1mm, and breath-hold at full inspiration. All images were de-identified before sending for analysis. This study was approved by the Institutional Review Board. Written informed consent was waived due to retrospective nature of the study.



2.2. Cost-Sensitive Uncertainty Hypergraph Learning

2.2.1. Pathological Features Initialization

In this stage, we extract the initial features from the patient’s CT images, consisting of regional features and radiomics features. We first apply the deep learning pre-trained method, named VB-Net (31), to segment the left/right lung, 5 lung lobes, 18 lung segments and infection lesions for each CT image in the portal software. In the expression of regional features, we generate a feature with a dimension of \mathbb{R}^{96} for each patient, expressing features such as the count of infected lesions and the mean value of lesion area. When extracting the radiomics features, we generated a feature with a dimension of \mathbb{R}^{93} for each patient, which means first-order intensity statistics and texture features. In the end, we concatenate the regional features and radiomics features obtained above to obtain an overall feature with a dimension of \mathbb{R}^C ($C = 191$) representing patient information.

2.2.2. Multi-Uncertainty Measurement

As shown in **Figure 2**, there are two types of uncertainty measurement in our method, namely “Epistemic Weighting” and “Aleatoric Weighting,” respectively. The epistemic uncertainty refers to the inability of model for classifying the lymph node involvement in patients with LUAD. We utilize the general Multilayer Perceptron (MLP) Neural Network with the dropout variation inference to classify the data based on the initialized features. Illustrated as the “Epistemic Weighting” module in **Figure 2**, denoted as $U^E \in \mathbb{R}^{N \times M}$, the effect of the dropout can be attributed to imposing a Gaussian distribution on each layer during the inference stage. For N samples, there are M layers that, respectively, generate the epistemic uncertainty weights in different levels. For the M layers, each case with \mathbf{x}^* features, is predicted for K times, the final epistemic weight for each case is calculated using the variance of these K values, formulated as Equation (1):

$$U_{q(\mathbf{y}^*|\mathbf{x}^*)}^E(\mathbf{y}^*) = \frac{1}{KM} \sum_{k=1}^K \sum_{l=1}^M \hat{\mathbf{y}}_{(l)}^*(\mathbf{x}^*, \omega^k) \quad (1)$$

where i denotes the i th sample and k denotes the k th test with dropout. (l) denote the l th layer of the constructed MLP model. $\hat{\mathbf{y}}^*$ denotes the corresponding output of the input \mathbf{x}^* . $\omega^k = \{\mathbf{W}_i\}^k$ denotes the trainable variables for a model at the k th time.

We adopt a statistical measuring method to generate the aleatoric uncertainty weights U^A . As shown in **Figure 2**, the dimension of aleatoric uncertainty weights is $\mathbb{R}^{N \times C}$, where N and C denotes the number of samples and the scale of features, respectively. For each feature, we estimate the weights of aleatoric uncertainty by minimizing the Kullback-Leibler (KL) divergence (32–34) between the standard feature distribution and the predicted features. The detailed theoretical derivation and demonstration of calculation can be found in UVHL (23) and the main formulation is following:

$$U^A(\mathbf{x}_i) = \sigma_{\Theta}^2(\mathbf{x}_i) = \exp(\alpha_{\Theta}(\mathbf{x}_i)) \quad (2)$$

where σ_{Θ}^2 denotes the predicted variance. To avoid the potential division by zero, $\alpha_{\Theta}(\mathbf{x})$ is the replacement of $\log \sigma_{\Theta}^2(\mathbf{x})$. Therefore, $\alpha_{\Theta} : \mathbb{R}^{191} \mapsto \mathbb{R}^1$ is the module to yield the aleatoric uncertainty score for each case.

2.2.3. Two-Stage Cost Sensitive Hypergraph Learning

To identify the LUAD cases with higher sensitivity, we design a two-stage cost sensitive hypergraph learning in the final step of our framework. Given the vertices with initialized pathological features as well as the corresponding two types of uncertainty weights, we sequentially construct the uncertainty-vertex hypergraph and conduct cost-sensitive hypergraph learning.

When constructing the uncertainty-vertex hypergraph, we take each vertex $v \in \mathcal{V}$ denoting one sample with the corresponding two types of uncertainty weights U^E and U^A . We use \mathcal{V} to denote the vertex set, \mathcal{E} denoting the hyperedges set, and \mathbf{W} denoting the pre-defined matrix of hyperedge weights. We adopt the k-nearest neighbors algorithm (KNN) to define the relationships for each vertex. There are two groups of hyperedges, respectively, stand for the regional features and radiomics features, denoted as \mathcal{E}_{reg} and \mathcal{E}_{rad} , which are represented by the corresponding incidence matrices $\mathbf{H}_{reg} \in \mathbb{R}^{N \times |\mathcal{E}_{reg}|}$ and $\mathbf{H}_{rad} \in \mathbb{R}^{N \times |\mathcal{E}_{rad}|}$. The final combined global incidence matrix $\mathbf{H} \in \mathbb{R}^{N \times (|\mathcal{E}_{reg}| + |\mathcal{E}_{rad}|)}$ can be formulated as Equation (3).

$$\mathbf{H}(v_i, e_j) = \begin{cases} U_i^A + U_i^E, & v_i \in e_j, e_j \in [\mathbf{H}_{reg} \parallel \mathbf{H}_{rad}] \\ 0, & v_i \notin e_j, e_j \in [\mathbf{H}_{reg} \parallel \mathbf{H}_{rad}] \end{cases} \quad (3)$$

where $[\parallel]$ denotes the concatenating operation between two matrices. Finish constructing the hyperedges, the uncertainty-vertex hypergraph can be denoted as $\mathcal{G} = \langle \mathcal{V}, \mathcal{E}, \mathbf{H}, \mathbf{W}, \mathbf{U} \rangle$, where \mathbf{U} is the summary matrix of U^A and U^E .

There are two stages of operating the cost sensitivity, namely latent cost sensitivity and supervising cost sensitivity. When measuring the epistemic uncertainty weights in the stage of “Multi-Uncertainty Measurement,” we design the first latent cost sensitivity by the modified cross-entropy loss function, formulated as follows:

$$\mathcal{L} = \frac{1}{N} \sum_i^N - [\lambda \cdot y_i \cdot \log(p_i) + (1 - \lambda) \cdot (1 - y_i) \cdot \log(1 - p_i)] \quad (4)$$

where y_i denotes the label of i th sample, whose value is 0 or 1 for negative case and positive case. $\lambda \in (0, 1)$ is the parameter to represent the degree of cost sensitivity, whose value larger the more sensitivity. The other cost sensitivity for supervising is designed in the procedure of hypergraph learning, formulated as:

$$Q_U(\mathbf{F}) = \arg \min_{\mathbf{F}} \{ \Omega(\mathbf{F}) + \psi \tilde{\mathcal{R}}_{emp}(\mathbf{F}) \} \quad (5)$$

where $\Omega(\cdot)$ and $\tilde{\mathcal{R}}_{emp}(\cdot)$ denote the smoothness regularizer function and the cost-sensitive empirical loss term, respectively. The hypergraph Laplacian matrix is

$\Theta_U = \mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^T \mathbf{D}_v^{-\frac{1}{2}}$. The smoothness regularizer function is formulated as:

$$\Omega(\mathbf{F}, \mathcal{V}, \mathbf{U}, \mathcal{E}, \mathbf{W}) = \text{tr}(\mathbf{F}^T (\mathbf{U}^T - \mathbf{U}^T \Theta_U \mathbf{U}) \mathbf{F}) \quad (6)$$

The cost-sensitive empirical loss term is designed as:

$$\begin{aligned} \tilde{\mathcal{R}}_{emp}(\mathbf{F}, \mathbf{U}) = & \sum_{k=1}^K \left[\lambda \|\mathbf{F}(:, k) - \mathbf{Y}(:, k)\|_{y=1}^2 \right. \\ & \left. + (1 - \lambda) \|\mathbf{F}(:, k) - \mathbf{Y}(:, k)\|_{y=0}^2 \right] \end{aligned} \quad (7)$$

where $\mathbf{F}(:, k)$ is the k_{th} column of \mathbf{F} . The λ in Equation (7) is same with the role in Equation (1).

The uncertainty vertex-weighted hypergraph loss function $\mathcal{R}_{emp}(\cdot)$ can be further rewritten as:

$$\tilde{\mathcal{R}}_{emp}(\mathbf{F}, \mathbf{U}, \lambda) = \lambda \triangleright \left[\text{tr}(\mathbf{F}^T \mathbf{U}^T \mathbf{U} \mathbf{F} + \mathbf{Y}^T \mathbf{U}^T \mathbf{U} \mathbf{Y} - 2\mathbf{F}^T \mathbf{U}^T \mathbf{U} \mathbf{Y}) \right] \quad (8)$$

Therefore, the target label matrix \mathbf{F} can be obtained as:

$$\mathbf{F} = \lambda \triangleright \left[\psi(\mathbf{U}^T - \mathbf{U}^T \Theta_U \mathbf{U} + \lambda \mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{U} \mathbf{Y} \right] \quad (9)$$

where \triangleright denotes the degree of cost-sensitive λ operating on the following item, referring the effect in Equations 1 and 7. With the generated label matrix $\mathbf{F} \in \mathbb{R}^{n \times K}$ ($K = 2$ in our task), the new coming testing case can be identified as LUAD or normal case accordingly.

3. EXPERIMENT

In this section, we will elaborate on the dataset, evaluation metrics, implementation, comparison methods, experimental results, and discussion.

3.1. Evaluation Metrics

In the experiment, we adopt six metrics to evaluate the accuracy of the model.

(1). Accuracy (*ACC*) represents the proportion of correct predictions by the model and can be calculated as $ACC = \frac{TP+TN}{TP+TN+FP+FN}$. (2). Sensitivity (*SEN*), (3). Specificity (*SPEC*), (4). Positive Predictive Value (*PPV*), and (5). Negative Predictive Value (*NPV*), respectively, represent the proportion of correct predictions among the positive sample values, negative sample values, positive predicted values, and negative predicted values. The calculation formulas can be found in **Table 1**. (6). Balance (*BAC*) represents the mean value of *SEN* and *SPEC*.

Sensitivity, known as true positive rate, represents the proportion of patients with lymph node involvement that are successfully detected in the task. Specificity represents the possibility of patients without lymph node involvement that are excluded. Ideally, the model with both high sensitivity and high specificity is what we most hope for, but in practice, there is a trade-off between these two indicators. Compared with specificity, higher sensitivity basically possesses greater practical value.

TABLE 1 | The definition of the confusion matrix for identification of lymph node involvement.

	Classify as lymph node involvement	Classify as non involvement	
Lymph node involvement	True Positive (<i>TP</i>)	False Negative (<i>FN</i>)	$SEN = \frac{TP}{TP+FN}$
Non involvement	False Positive (<i>FP</i>)	True Negative (<i>TN</i>)	$SPEC = \frac{TN}{TN+FP}$
	$PPV = \frac{TP}{TP+FP}$	$NPV = \frac{TN}{TN+FN}$	

3.2. Implementation

The entire dataset contains 61 CT images, of which 26 are lymph node involvement and the remaining 35 are on the contrary, are randomly partitioned into 10 subsets when comparing our model with the comparison models. In the task, the cross-validation process is performed 10 times, each time a subset is selected as the validation set, and the rest as the training set. To reduce the impact of random data on the results, the value of each metric in the experiment is an average of 10 times, and the standard deviation is reported as a comparison. To prevent inductive bias, each dimension of the training set features is normalized to [0, 1] using its own mean and variance and samples in validation set utilize the same parameters to normalize.

The uncertainty score \mathcal{U}_i of each sample and the uncertainty measurement model are generated by the overall training set. The algorithm used to construct the incidence matrix of hypergraph is K -nearest neighbors (KNN), which leads the choice of parameter \mathcal{K} to affect the effect. However, choosing \mathcal{K} is not a easy job. A hyperedge from a large \mathcal{K} connects too many nodes, which may over-describe the relationship between the data and generate noise. On the contrary, a hyperedge with small \mathcal{K} means that the number of connected nodes is small, which limits the exploration ability of the high-order relationship of hypergraph and not obtain full information. We conduct a strategy to learn the proper parameter \mathcal{K} automatically here. We put 2 to 20 in the candidate pool of \mathcal{K} to select the applicable \mathcal{K} for the task. In one training and testing, we cross-validate the training data 10 times for each \mathcal{K} in the candidate pool to obtain prediction values on different \mathcal{K} . The \mathcal{K} with the highest predicted score will be used in testing, so the whole process is the automatic selection of \mathcal{K} . The total training time for each fold is about 1 h, while the testing time is extremely fast, only taking about 5 s.

3.3. Comparison Methods

The following methods are compared by our experiment:

1. Support Vector Machine (SVM) (35): It is a linear classifier that uses supervised learning to perform two-class classification of data. It relies on the convex quadratic programming problem to separate the samples correctly and away from the classification hyperplane.
2. Transductive Hypergraph Learning (tHL) (36): It is an algorithm for hypergraph embedding and transduction inference, mainly extending the spectral clustering technology of graphs to hypergraphs.

- Multilayer Perceptron (MLP) Neural Network (37, 38): It contains a multi-layer feedforward neural network that maps multiple inputs to a single output.
- GNN (39): It is a basic graph neural network that applies the local first-order approximation of the spectral graph convolution to determine the convolutional network structure for semi-supervised classification.
- UVHL (23): It is an uncertainty vertex-weighted hypergraph learning method, which can reduce the problems caused by noisy data and confusing cases with clinical or imaging features.
- B-GNN (40): It is a binary graph convolutional network, in which some floating-point operations are replaced by binary operations to achieve inference acceleration. A back-propagation method based on gradient approximation is used to train the binarized graph convolutional layer.

- BC-GNN (40): It is an improved version of B-GNN, adding a cost-sensitive loss function.

3.4. Experimental Results

The experimental mean results and phenomena of our model and compared models can be seen in **Figure 3**; **Table 2**, and the following can be observed:

- Comparing all methods, our proposed model is in a leading position in various metrics. Compared with non-graph-based methods (SVM and MLP), our models has a great lead. There are 12.04 and 38.89% improvements in ACC metric for the two, respectively, which shows that hypergraph has the ability to describe the correlation and handle this task.
- In the GNN-based methods, compared with GNN, B-GNN, BC-GNN in ACC, the improvement is 24.22, 5.26, and 3.90%,

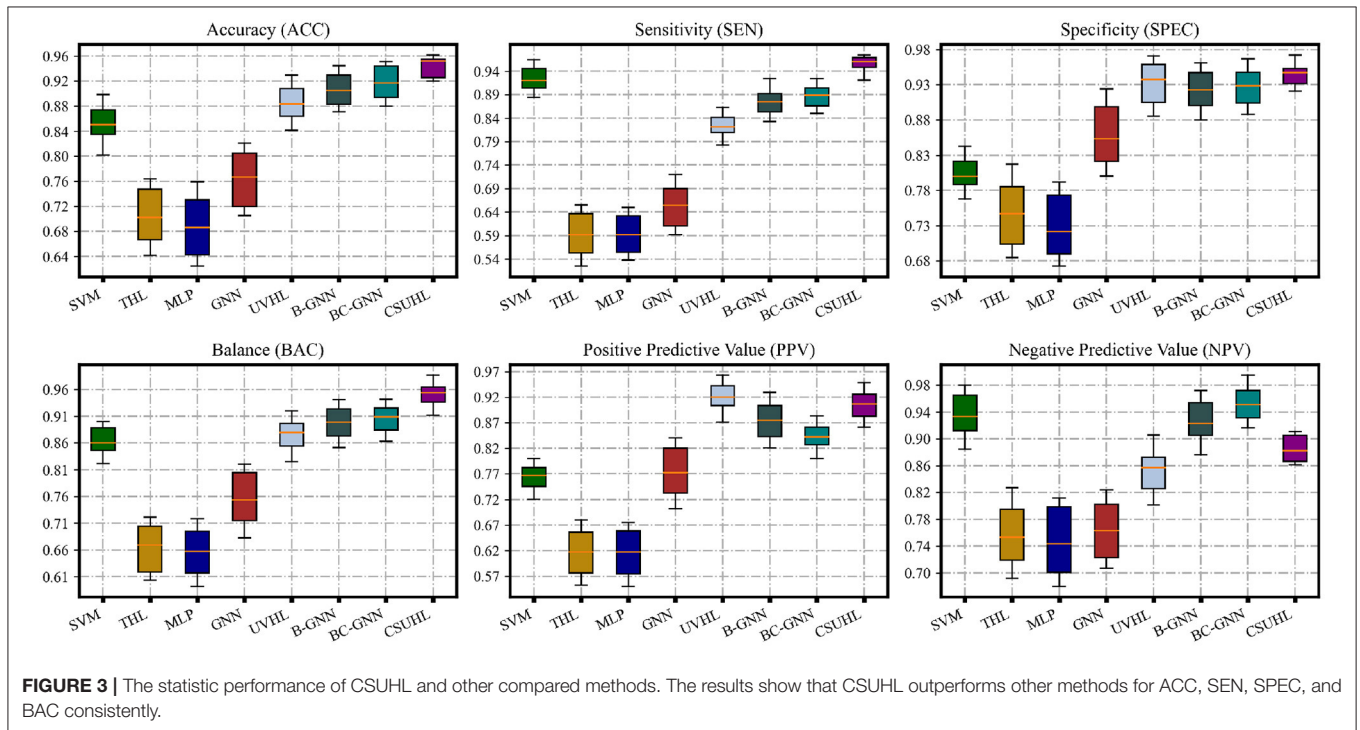


FIGURE 3 | The statistic performance of CSUHL and other compared methods. The results show that CSUHL outperforms other methods for ACC, SEN, SPEC, and BAC consistently.

TABLE 2 | Prediction accuracy comparison of different methods on our collected LUAD dataset.

Methods		ACC		SEN		SPEC		BAC		PPV		NPV	
SVM	(<i>p</i> -value)	0.85000	2.324e-5	0.92000	5.624e-4	0.80000	1.824e-4	0.86000	6.815e-5	0.76667	0.0498e-5	0.93333	6.781e-5
THL	(<i>p</i> -value)	0.70238	1.173e-5	0.59167	1.438e-6	0.74667	4.235e-4	0.66917	1.037e-4	0.61667	0.1.237e-5	0.75333	3.283e-6
MLP	(<i>p</i> -value)	0.68571	6.734e-4	0.59167	3.568e-5	0.72167	8.967e-3	0.65667	2.358e-4	0.61667	8.845e-4	0.74333	2.781e-5
GNN	(<i>p</i> -value)	0.76667	4.891e-4	0.65385	6.784e-4	0.85294	3.578e-4	0.75339	3.567e-4	0.77273	9.487e-4	0.76316	7.034e-4
UVHL	(<i>p</i> -value)	0.88333	2.346e-3	0.82143	7.624e-4	0.93750	6.78e-4	0.87946	1.895e-3	0.92000	-	0.85714	3.181e-4
B-GNN	(<i>p</i> -value)	0.90480	7.823e-3	0.87500	2.135e-3	0.92300	7.895e-3	0.89900	8.233e-3	0.87500	9.356e-3	0.92300	9.392e-3
BC-GNN	(<i>p</i> -value)	0.91667	4.721e-2	0.88889	1.468e-3	0.92857	2.568e-2	0.90873	9.134e-3	0.84211	7.804e-3	0.95122	-
CSUHL	(<i>std</i>)	0.95238†	±0.0346	0.96000†	±0.0596	0.94737†	±0.0277	0.95368†	±0.0150	0.90654	±0.0286	0.88235	±0.0735

For each 10-fold, we compute the accuracy of the proposed method on testing data, and compare them with those of CSUHL via paired t-test to generate the *p*-values for each metric. ("†" denotes significance level is reached as *p* - value < 0.05). The bold values represent the best values of the indicators in each set of experiments.

- respectively, which proves that our methods can describe complex associations better.
- Compared with hypergraph-based methods such as THL and UVHL, the model has 35.59 and 7.82% rises on the acc, respectively, which benefits from the uncertainty measurement and multiple loss.
 - Except for ACC, our method is the only one that exceeds and close to 95% on SEN and SPEC, respectively, which has practical value for actual medical diagnosis.
 - From **Figure 3**, it can be observed from the standard deviation that our model has more stable results compared to other comparison methods, which shows that our model provides more reliable and robust prediction results.

4. DISCUSSION

To evaluate the effectiveness of different uncertainty and different hypergraphs and different cost sensitivity, in this section, we conduct ablation experiments, respectively, to determine the contribution of each component.

4.1. Study on Multi-Uncertainty

To evaluate the effectiveness of different uncertainty, we conduct an ablation study, which uses aleatoric uncertainty or epistemic uncertainty, respectively.

4.1.1. Aleatoric Uncertainty

Denoted as \mathcal{U}^A , the results of using aleatoric uncertainty individually are shown in row 1 of **Table 3**.

We can find out that the use of both types of uncertainty brings about 7.58, 4, and 10.53% growth in ACC, SEN, and SPEC, respectively, than using aleatoric uncertainty. It is worth mentioning that even with only aleatoric uncertainty, the model is still higher than most comparison methods in ACC and better than other methods in SEN according to **Table 2**.

More specifically, the pathological features with higher aleatoric uncertainty weights are consistent with the clinical experience, such as the distribution of different nodules, *i.e.*, lobulated nodules, spiculate nodules, and globular nodules.

4.1.2. Epistemic Uncertainty

Denoted as \mathcal{U}^E , the results of using epistemic uncertainty individually are shown in row 2 of **Table 3**.

It can be observed using epistemic uncertainty alone lags behind 10.49, 7.85, and 12.54 than CSUHL in acc, sen, and spec metrics, respectively. Compared with the results of using aleatoric uncertainty, the indicators using epistemic uncertainty are lower, indicating that aleatoric uncertainty plays a greater role in our model. The combination of the two uncertainties is better than the single-use, which proves the effectiveness of multi-uncertainty.

4.2. Study on Types of Hypergraph

To evaluate the effectiveness of different hypergraphs, we conduct an ablation study, using regional hypergraph or radiomics hypergraph, which use regional features and radiomics features from CT, respectively.

4.2.1. Regional Hypergraph

Denoted as \mathcal{G}_{reg} , the results of using regional hypergraph individually are shown in row 4 of **Table 3**.

The regional hypergraph only has an accuracy rate of about 80%, and the same for SEN and SPEC, indicating that only extracting the regional features of CT has little effect and the regional hypergraph cannot provide accurate correlation information.

4.2.2. Radiomics Hypergraph

Denoted as \mathcal{G}_{rad} , the results of using radiomics hypergraph individually are shown in row 5 of **Table 3**.

The results of radiomics hypergraph are much better than the former, with ACC and SEN exceeding 90%, although there is still a gap in the combination of two hypergraphs. It can be found that the results in radiomics hypergraph have better sensitivity than specificity, which proves that the radiomic hypergraph has more advantages in identifying lymph node involvement. The combined hypergraph is higher in all indicators than when used alone, showing that it has the ability to utilize a variety of different features.

TABLE 3 | Prediction accuracy comparison of different methods on our collected LUAD dataset.

Methods		ACC	SEN	SPEC	BAC	PPV	NPV
1) Aleatoric Uncertainty (\mathcal{U}^A)	(std)	0.88525 ±0.1845	0.92308 ±0.2451	0.85714 ±0.1684	0.89011 ±0.1795	0.82759 ±0.0781	0.93750 ±0.1864
2) Epistemic Uncertainty (\mathcal{U}^E)	(std)	0.85246 ±0.0351	0.88462 ±0.0763	0.82857 ±0.0374	0.85659 ±0.0746	0.79310 ±0.0890	0.90625 ±0.0785
3) CSUHL ($\mathcal{U}^A + \mathcal{U}^E$)	(std)	0.95238† ±0.0346	0.96000† ±0.0596	0.94737† ±0.0277	0.95368† ±0.0150	0.90654 ±0.0286	0.88235 ±0.0735
4) Regional Hypergraph (\mathcal{G}_{reg})	(std)	0.80328 ±0.0567	0.80769 ±0.978	0.80000 ±0.1643	0.80385 ±0.0776	0.75000 ±0.1347	0.84848 ±0.1613
5) Radiomics Hypergraph (\mathcal{G}_{rad})	(std)	0.90164 ±0.0891	0.92308 ±0.0346	0.88571 ±0.0917	0.90440 ±0.0176	0.85714 ±0.0783	0.93939 ±0.0635
6) CSUHL ($\mathcal{G}_{reg} + \mathcal{G}_{rad}$)	(std)	0.95238† ±0.0346	0.96000† ±0.0596	0.94737† ±0.0277	0.95368† ±0.0150	0.90654 ±0.0286	0.88235 ±0.0735
7) Latent Cost Sensitivity	(std)	0.86885 ±0.0678	0.88462 ±0.0341	0.85714 ±0.0867	0.87088 ±0.1456	0.82143 ±0.1034	0.90909 ±0.0918
8) Supervising Cost Sensitivity	(std)	0.91803 ±0.0451	0.92308 ±0.0813	0.91429 ±0.0561	0.91868 ±0.0971	0.88889 ±0.0936	0.94118 ±0.0771
9) CSUHL	(std)	0.95238† ±0.0346	0.96000† ±0.0596	0.94737† ±0.0277	0.95368† ±0.0150	0.90654 ±0.0286	0.88235 ±0.0735

For each 10-fold, we compute and report the average performance of the proposed method on testing data. The bold values represent the best values of the indicators in each set of experiments.

4.3. Study on Cost Sensitivity

To evaluate the effectiveness of different cost sensitivity, we conduct an ablation study, which uses latent cost sensitivity or supervising cost sensitivity, respectively.

4.3.1. Latent Cost Sensitivity

The results of using latent cost sensitivity individually are shown in row 7 of **Table 3**.

When only latent cost sensitivity is used, it is equivalent to not using supervising cost sensitivity. As a result, the hypergraph information cannot be captured in order that various indicators are significantly reduced.

4.3.2. Supervising Cost Sensitivity

The results of using supervising cost sensitivity individually are shown in row 8 of **Table 3**.

Compared with the combination of the two cost sensitivity, the supervising cost sensitivity has an accuracy disparity of about 3.61%, but it is higher than latent cost sensitivity. It should be noticed that cost sensitivity of using only supervising is the highest among the three on NPV, indicating that the true label is mostly negative in the samples identified as non lymph node involvement. In general, using two cost sensitivities together is better than using one of them, proving the effectiveness of cost sensitivity component.

5. CONCLUSION

In this paper, we propose a cost-Sensitive Uncertainty Hypergraph Learning (CSUHL) to identify lung adenocarcinoma (LUAD) cases with lymph node (LN) metastasis from the cases without lymph node (LN) metastasis. Confronting the challenging issues from the shortage of high-quality data and unreliable sensitivity of diagnosis, our proposed method employs three stages, namely “Pathological Features Initialization,” “Multi-Uncertainty Measurement,” and “Two-Stage Cost Sensitive Hypergraph Learning” to represent the

complex clinical information and formulate the high-order data correlation among the known LUAD with LN metastasis cases and the LUAD without LN metastasis cases. Through the epistemic and aleatoric uncertainty as well as the two types of cost sensitivity (latent and supervising), our method is capable of outperforming state-of-the-art methods on our collected LUAD dataset across the board.

In future work, we will further investigate the practical limitations on the computer-aid-diagnosis (CAD), such as enhancing the speed of inference, transferring the model to learn, and predicting the other related downstream clinical tasks.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Clinical Research Ethics Committee of China-Japan Friendship Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

QM and DD: study design. QM, QY, YZ, and CL: data collection. JY, JZ, and DD: data analysis. JZ and DD: supervision. JY and DD: manuscript writing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Peking science and technology fund (No. Z191100006619008) and Elite Medical Professionals project of China-Japan Friendship Hospital (No. ZRJY2021-GG02).

REFERENCES

- Planchard D, Popat S, Kerr K, Novello S, Smit E, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* (2018) 29:iv192–iv237. doi: 10.1093/annonc/mdy275
- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2021) 71:209–49. doi: 10.3322/caac.21660
- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* (2018) 68:394–424. doi: 10.3322/caac.21492
- Travis WD. Pathology of lung cancer. *Clin Chest Med.* (2011) 32:669–92. doi: 10.1016/j.ccm.2011.08.005
- Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. *Trends Mol Med.* (2019) 25:585–94. doi: 10.1016/j.molmed.2019.04.012
- Zhang L, Zhang Z, Yu Z. Identification of a novel glycolysis-related gene signature for predicting metastasis and survival in patients with lung adenocarcinoma. *J Transl Med.* (2019) 17:1–13. doi: 10.1186/s12967-019-02173-2
- Vansteenkiste JF, Stroobants SG, De Leyn P, Dupont PJ, Bogaert J, Maes A, et al. Lymph node staging in non-small-cell lung cancer with FDG-PET scan: a prospective study on 690 lymph node stations from 68 patients. *J Clin Oncol.* (1998) 16:2142–9.
- De Leyn P, Lardinois D, Van Schil PE, Rami-Porta R, Passlick B, Zielinski M, et al. ESTS guidelines for preoperative lymph node staging for non-small cell lung cancer. *Elsevier Science BV.* (2007) 32:1–8. doi: 10.1016/j.ejcts.2007.01.075
- Chen D, Ding Q, Wang W, Wang X, Wu X, Mao Y, et al. Characterization of extracapsular lymph node involvement and its clinicopathological characteristics in stage II–III lung adenocarcinoma. *Ann Surg Oncol.* (2021) 28:2088–98. doi: 10.1245/s10434-020-09154-6
- Park JK, Kim JJ, Moon SW, Lee KY. Lymph node involvement according to lung adenocarcinoma subtypes: lymph node involvement is influenced by lung adenocarcinoma subtypes. *J Thor Dis.* (2017) 9:3903. doi: 10.21037/jtd.2017.08.132
- Dietel M, Bubendorf L, Dingemans AMC, Doooms C, Elmberger G, Garcia RC, et al. Diagnostic procedures for non-small-cell lung cancer (NSCLC):

- recommendations of the European Expert Group. *Thorax*. (2016) 71:177–84. doi: 10.1136/thoraxjnl-2014-206677
12. León-Atance P, Moreno-Mata N, González-Aragoneses F, Cañizares-Carretero MÁ, García-Jiménez MD, Genovés-Crespo M, et al. Multicenter analysis of survival and prognostic factors in pathologic stage I non-small-cell lung cancer according to the new 2009 TNM classification. *Archivos de Bronconeumología (English Edition)*. (2011) 47:441–6. doi: 10.1016/j.arbres.2011.04.004
 13. Choe J, Lee SM, Do KH, Lee G, Lee JG, Lee SM, et al. Deep learning-based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*. (2019) 292:365–73. doi: 10.1148/radiol.2019181960
 14. Palumbo B, Bianconi F, Palumbo I, Fravolini ML, Ministrini M, Nuvoli S, et al. Value of shape and texture features from 18F-FDG PET/CT to discriminate between benign and malignant solitary pulmonary nodules: an experimental evaluation. *Diagnostics*. (2020) 10:696. doi: 10.3390/diagnostics10090696
 15. Zhong Y, Yuan M, Zhang T, Zhang YD, Li H, Yu TF. Radiomics approach to prediction of occult mediastinal lymph node metastasis of lung adenocarcinoma. *Am J Roentgenol*. (2018) 211:109–113. doi: 10.2214/AJR.17.19074
 16. Dai C, Xie H, Kadeer X, Su H, Xie D, Ren Y, et al. Relationship of lymph node micrometastasis and micropapillary component and their joint influence on prognosis of patients with stage I lung adenocarcinoma. *Am J Surg Pathol*. (2017) 41:1212–20. doi: 10.1097/PAS.0000000000000901
 17. Shim SS, Lee KS, Kim BT, Chung MJ, Lee EJ, Han J, et al. Non-small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology*. (2005) 236:1011–9. doi: 10.1148/radiol.2363041310
 18. Yu W, Tang C, Hobbs BP, Li X, Koay EJ, Wistuba II, et al. Development and validation of a predictive radiomics model for clinical outcomes in stage I non-small cell lung cancer. *Int J Radiat Oncol Biol Phys*. (2018) 102:1090–7. doi: 10.1016/j.ijrobp.2017.10.046
 19. Tsutani Y, Miyata Y, Nakayama H, Okumura S, Adachi S, Yoshimura M, et al. Prediction of pathologic node-negative clinical stage IA lung adenocarcinoma for optimal candidates undergoing sublobar resection. *J Thor Cardiovas Surg*. (2012) 144:1365–71. doi: 10.1016/j.jtcvs.2012.07.012
 20. Higaki T, Nakamura Y, Tatsugami F, Nakaura T, Awai K. Improvement of image quality at CT and MRI using deep learning. *Japanese J Radiol*. (2019) 37:73–80. doi: 10.1007/s11604-018-0796-2
 21. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Trans Med Imag*. (2018) 37:1370–81. doi: 10.1109/TMI.2018.2823083
 22. Tan W, Liu P, Li X, Liu Y, Zhou Q, Chen C, et al. Classification of COVID-19 pneumonia from chest CT images based on reconstructed super-resolution images and VGG neural network. *Health InfSci Syst*. (2021) 9:1–12. doi: 10.1007/s13755-021-00140-0
 23. Di D, Shi F, Yan F, Xia L, Mo Z, Ding Z, et al. Hypergraph learning for identification of COVID-19 with CT imaging. *Med Image Anal*. (2021) 68:101910. doi: 10.1016/j.media.2020.101910
 24. Feng Y, You H, Zhang Z, Ji R, Gao Y. Hypergraph neural networks. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33 Honolulu, HI (2019). p. 3558–65.
 25. Fan H, Zhang F, Wei Y, Li Z, Zou C, Gao Y, et al. Heterogeneous hypergraph variational autoencoder for link prediction. *IEEE Trans Pattern Anal Mach Intell*. (2021). doi: 10.1109/TPAMI.2021.3059313. [Epub ahead of print].
 26. Gao Y, Zhang Z, Lin H, Zhao X, Du S, Zou C. Hypergraph learning: methods and practices. *IEEE Trans Pattern Anal Mach Intell*. (2020). doi: 10.1109/TPAMI.2020.3039374
 27. Di D, Li S, Zhang J, Gao Y. Ranking-based survival prediction on histopathological whole-slide images. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima: Springer (2020). p. 428–38.
 28. Ruan D, Ji S, Yan C, Zhu J, Zhao X, Yang Y, et al. Exploring complex and heterogeneous correlations on hypergraph for the prediction of drug-target interactions. *Patterns*. (2021) 2:100390. doi: 10.1016/j.patter.2021.100390
 29. Zhou CM, Wang Y, Ye HT, Yan S, Ji M, Liu P, et al. Machine learning predicts lymph node metastasis of poorly differentiated-type intramucosal gastric cancer. *Sci Rep*. (2021) 11:1–7. doi: 10.1038/s41598-020-80582-w
 30. Wu Y, Rao K, Liu J, Han C, Gong L, Chong Y, et al. Machine learning algorithms for the prediction of central lymph node metastasis in patients with papillary thyroid cancer. *Front. Endocrinol*. (2020) 11:816. doi: 10.3389/fendo.2020.577537
 31. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Lung infection quantification of COVID-19 in CT images with deep learning. *arXiv preprint arXiv:200304655*. (2020). doi: 10.1002/mp.14609
 32. Van Erven T, Harremoës P. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans Inf Theory*. (2014) 60:3797–820. doi: 10.1109/TIT.2014.2320500
 33. Hershey JR, Olsen PA. Approximating the Kullback-Leibler divergence between Gaussian mixture models. In: *IEEE International Conference on Acoustics*. Vol. 4 (2007). Honolulu, HI p. IV–317.
 34. Moreno PJ, Ho PP, Vasconcelos N. A Kullback-Leibler divergence based kernel for SVM classification in multimedia applications. In: *Advances in Neural Information Processing Systems*. (2004). p. 1385–1392.
 35. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. (1995) 20:273–97.
 36. Zhou D, Huang J, Schölkopf B. Learning with hypergraphs: clustering, classification, and embedding. In: *Advances in Neural Information Processing Systems*. Citeseer (2007). p. 1601–1608.
 37. Thimm G, Fiesler E. High-order and multilayer perceptron initialization. *IEEE Trans Neural Netw*. (1997) 8:349–59.
 38. Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Exp Syst Appl*. (2011) 38:13475–81. doi: 10.1016/j.eswa.2011.04.149
 39. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:160902907*. (2016).
 40. Wang J, Wang Y, Yang Z, Yang L, Guo Y. Bi-gcn: Binary graph convolutional network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021). p. 1561–70.

Conflict of Interest: JZ is employed by Tencent AI Lab, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Ma, Yan, Zhang, Yu, Zhao, Liang and Di. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.